

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 22, 2003, 13:48:51 ; Search time 160 Seconds

(without alignments)  
5009.446 Million cell updates/sec

Title: US-09-745-506-74

Perfect score: 2789  
Sequence: 1 GTGATGTCTTACTTGCTGCT.....TCTGTTTACTTACATTCAC 1553

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame\_n2p.model -DEV=xlp  
-Q/cgn2.1/USPTO/US09745506/runatc\_22082003\_132710.10741/apc\_query.fasta.1.1735  
-DB=SPREMBL\_23 -QMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=dits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cd1  
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pic -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09745506.ecgn.1.1.223.ernatc\_22082003\_132710.10741 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEBUFFER -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SPREMBL\_23:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phase:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_virus:\*  
17: sp\_bacterioplasmid:\*  
17: sp\_archaealplasmid:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1373	49.2	340	11	Q8C608 mus musculu

2	707	25.3	159	11	Q9D146	Q9D146 mus musculu
3	358.5	12.9	366	16	Q8A972	Q8A972 thermomane
4	332	11.2	376	16	Q8CSD9	Q8CSD9 staphylococ
5	309.5	11.1	372	16	Q8EP20	Q8EP20 oceanobacti
6	279	10.0	366	16	Q8NMB9	Q8NMB9 staphylococ
7	271.5	9.7	258	16	Q8RE15	Q8RE15 fusobacteri
8	256	9.2	242	5	Q8SV52	Q8SV52 encephalit
9	246.5	8.8	433	16	Q8PKN9	Q8PKN9 corynebacte
10	230	8.2	380	16	Q8NNG6	Q8NNG6 corynebacte
11	203.5	7.3	265	16	Q8DNV4	Q8DNV4 streptococ
12	203.5	7.3	309	16	Q8G5M5	Q8G5M5 bifidobacte
13	189.5	6.8	262	16	Q8DT99	Q8DT99 streptococ
14	185.5	6.7	262	2	P95777	P95777 streptococ
15	181.5	6.5	265	2	Q914G2	Q914G2 lactobacilli
16	174	6.2	262	16	Q8DZAT	Q8DZAT streptococ
17	173	6.2	262	16	Q8E4W8	Q8E4W8 streptococ
18	151.5	5.4	298	17	Q8TMA2	Q8TMA2 methanosa
19	143.5	5.1	253	16	Q8EUT6	Q8EUT6 mycoplasma
20	138	4.9	306	17	Q8PTJ7	Q8PTJ7 methanosa
21	135.5	4.9	248	17	Q8TW44	Q8TW44 methanopyru
22	118.5	4.3	1637	6	Q9XSV8	Q9XSV8 bos taurus
23	118.5	4.3	5146	6	Q8SPM4	Q8SPM4 bos taurus
24	114	4.1	250	16	Q8EDX0	Q8EDX0 shewanella
25	114	4.1	1260	5	Q9XZC5	Q9XZC5 cryptospori
26	112	4.0	251	2	Q8KTX3	Q8KTX3 vibrio fisc
27	109	3.9	249	16	Q8D2D0	Q8D2D0 wigleswort
28	106.5	3.8	556	16	Q8RAG1	Q8RAG1 thermomane
29	106	3.8	3352	2	Q9JH83	Q9JH83 streptomyc
30	105.5	3.8	252	17	Q8U1Y9	Q8U1Y9 pyrococcus
31	105.5	3.8	566	16	Q9WZU1	Q9WZU1 thermotoga
32	105	3.8	644	10	Q38736	Q38736 antirrhinum
33	104.5	3.7	456	16	Q8ECAB	Q8ECAB shewanella
34	104.5	3.7	1052	2	Q93CZ8	Q93CZ8 streptomyc
35	103	3.7	653	4	Q92815	Q92815 homo sapien
36	102.5	3.7	4123	4	Q75851	Q75851 homo sapien
37	102	3.6	554	16	Q8ELV1	Q8ELV1 shewanella
38	101.5	3.6	850	16	Q8DLR3	Q8DLR3 synechococ
39	101	3.6	238	1	P94953	P94953 methanopyru
40	101	3.6	253	17	Q8RZC0	Q8RZC0 methanopyru
41	101	3.7	488	11	Q8BGE7	Q8BGE7 mus musculu
42	101	3.6	597	5	Q9VV09	Q9VV09 drosofila
43	101	3.6	804	16	Q8RF68	Q8RF68 escherichia
44	100	3.6	221	1	Q50523	Q50523 methanobact
45	100	3.6	804	16	Q8XCZ5	Q8XCZ5 escherichia

## ALIGNMENTS

### RESULT 1

ID Q8C608 PRELIMINARY; PRT; 340 AA.

AC Q8C608: 01-MAR-2003 (TREMUREL. 23, Created)  
DT 01-MAR-2003 (TREMUREL. 23, Last sequence update)  
DT 01-MAR-2003 (TREMUREL. 23, Last annotation update)  
DE N9gl interacting factor 3-like 1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=22354683; PubMed=1246681;  
RA The FANTOM Consortium.  
RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK076788; BAC36481.1; -  
SQ SEQUENCE 340 AA; 37822 MW; 82DC63DB90B68B4E CRC64;

Alignment Scores:

Pred. No.:	3,856,115	Length:	340
Score:	1373.00	Matches:	25
Percent Similarity:	90.57%	Conservative:	23
Best local Similarity:	83.33%	Mismatches:	30
Query Match:	49.23%	Indels:	0
DB:	11	Gaps:	0

```

OY 1118 ATGTCCATCATGATCTTGTGATGCTGTCCCAAGATTAATGTCACCTCTGGAA 1177
DB 101 MeSerHisHisAspValLeuAspAlaHisSerLysGlyLeuValHisLeuGlu 120
OY 1178 CACAGACACGACGAGGAGGCTTTCTTGTGACCTTCGAGATATGCTGATTCCTACTTG 1237
DB 121 HisSerAspThrGluArgGlyPheLeuSerGluLeuGlnGluMetLeuGlyValHisPhe 140
OY 1238 GACATATAGATTAATATATATCATCATGACGACGACGACCTCTCTGAGGCTTA 1294
DB 141 GluAspHisLeuHisLeuSerGluThrAspArgAspProLeuArgValVal 159

RESULT 3
OY 08R972 PRELIMINARY; PRT; 376 AA.
AC 08R972:
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein TRe1753.
GN TRe1753.
OS Thermoaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoaerobacteriales;
OC Thermoaerobacteriaceae; Thermoaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL: AEO13129; AAA24947.1;
DR InterPro: IPR002678; DUF34.
DR Pfam: PF01784; DUF34; 1.
DR TrEMBL: TIGR00486; TIGR00486; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 376 AA; 41674 MW; 557753F2A1F245C6 CRC64;

Alignment Scores:
Pred. No.: 1.91e-23 Length: 376
Score: 358.50 Matches: 105
Percent Similarity: 48.47% Conservative: 85
Best Local Similarity: 26.79% Mismatches: 123
Query Match: 12.85% Indels: 79
DB: 16 Gaps: 14

US-09-745-506-74 (1-1553) x 08R972 (1-376)
OY 254 AAGGCTCTCTTCTTCCCTGGAATGACTTTCATCCCTCGTTTGTGAGAGTTGGAC 313
DB 11 GluThrIleValSerIleMetAspArgLeuAlaProArgLysPheAlaGlnAspTrpAsp 30
OY 314 AATGTTGATTTAGTGTGGAACCAAGCCACATATCATTAATACACTTCTCTGACC 373
DB 31 AsnValGlyLeuIleVal---GlyAspProLeuGlnAspValSerLysIleLeuValAla 49
OY 374 AATGACCTGACGTAGAGATGATGAGAGAGGCTGCTGCAAAAAGAGACCTCATTTCTC 433
DB 50 LeuAspAlaThrHisGluValValGluGluAlaIleGluLysValAspMetIleVal 69
OY 434 TCCATACACCGGCTATCTTCGAGCCATGAGGACGATACCTGAGGAACATGGAAGAG 493
DB 70 ThrHisHisProLeuIlePheLysSerIleLysSerValArgSerAspTrpProValGly 89
OY 494 CGCGCTGATCCGGGCTGTGGAACAGAGTGATCTACTCTCTCATAGAGCTAT 553
DB 90 MetLeuValMetLysLeuIleLysHisGluIleSerLeuValAlaHisThrSerPhe 109
OY 554 GATGCTGCGCCCGGAGGCGCTCAACAACATGCTGCTTAAGGCTTGAGCTTATCCTCC 613
DB 111 111111 1111111111 111111

```

```

DB 110 AspIleAlaProAsnGlnMetAsnAspIleLeuGlnHisLeuGlyLeu----- 126
OY 614 AGGCCATACATCTCTTCCAAAGCTCCCACTACCTTACAGAGGAAACCCAGATGAA 673
DB 127 -----TyrGluAspGluVal 131
OY 674 TTCACGTTATATACACCAAGACCTGACAAAGTCATGCTGCAGTG---AAAGAAAT 730
DB 132 LeuAspValThrTyrSerGluAspTyrLysValValValTyrValProLysGlyTyr 151
OY 731 GACGCTGTT-----TCT 742
DB 152 GluAspValValArgAsnAlaMetCysSerAlaGlyAlaGlyPheIleGlyAsnTyrSer 171
OY 743 GTCACTCTTTTCTGCTAG---ACTGCT----- 769
DB 172 AspSerThrPheGlnValGluGlyThrGlyThrPheLysProMetGluGlyAlaAspPro 191
OY 770 -----AATGAGAACAAACAGGATTAATCTGAATGT 802
DB 192 PheIleGlyGluValGlyLysLeuGluLysValGluGluIleArgLeuGluThrValVal 211
OY 803 ACTCAGAGGCTTTGATGACAGGTGTGATTTCTTCCCG-----AACAA 850
DB 212 ProGlnTyrLeuAspLysValIleAsnAlaMetLeuArgValHisProTyrGluGlu 231
OY 851 CAACTTATATCAGAACGGAATTTCTCACTGAGAAAGCTTGTCTTACATCTGGA 910
DB 232 ValAlaTyrAspValTyrSerLeuAlaAsnLeuArgIle-----GluTyrGly 247
OY 911 ATGCGAGGTTATGACACATGATGATGATGCTGCTGCTGCAACATGATGATGATA 970
DB 248 LeuGlyArgIleGlyValIleGluGluThr---ThrLeuGluGluAlaLeuVal 266
OY 971 AAAAGACCTAAATATCTCATATTCCTTACCTTGGGCTGGGAGAACCTTAG 1030
DB 267 LysAlaLysLeuLysAlaGluSerLeuVal-----ValGly---AspLeuGln 282
OY 1031 TTCACGATCAAGACGCGGCGCTGCTGCTGCTGCTGCGAGAC 1075
DB 283 ArgArgValLysLysValAlaValAlaCysGlySerIleValSerLeuIleHisLysAla 302
OY 1076 GTTCTGAGGCTGTGAGGCTGACCTTATCTTACCTCAGAGTGAAGATGCTCATGATACT 1135
DB 303 ValMetLysGly-----AlaAspValLeuIleThrAlaAspIleGlyTyrHisAspAla 320
OY 1136 TTGATCTCTGCTTCCCAAGATTAATGATCATCTCTGTGACACAGCAACATGAAAGA 1195
DB 321 ValGluAlaGlnHisLeuGlyLeuAlaLeuIleAspAlaGlyHisPheAlaThrGluAsn 340
OY 1196 GCGTTTCTTCTGACCTTCGAGATATGCTGATTCCTCATTTGGAG-----AATAG 1246
DB 341 IleAlaValArgPheIleAlaGluTyrLeuIleAspGluThrGlnLysGlnGlyHisGlu 360
OY 1247 ATTAATATATCTTATCAGAGACTGACAGGAGACCT 1282
DB 361 IleGluValPheValSerGluValGlnLysAspPro 372

RESULT 4
OY 08CSD9 PRELIMINARY; PRT; 366 AA.
AC 08CSD9:
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SE1246.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;

```

RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
 RA Chen Z., Wen Y.,  
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AE016748; AA004845.1; -  
 KW Hypothetical protein; complete proteome.  
 SQ SEQUENCE 366 AA; 41639 MW; BF3A88EE77462A1 CRC64;

## Alignment Scores:

Pred. No.:	3,02e-19	Length:	366
Score:	312.00	Matches:	100
Percent Similarity:	44.78%	Conservative:	80
Best Local Similarity:	24.88%	Mismatches:	130
Query Match:	11.19%	Indels:	92
DB:	16	Gaps:	16

US-09-745-506-74 (1-1553) x Q8CSD9 (1-366)

```

QY 245 ATGATTTGAAAGGCTCTCTCTCTCTCTGAAATGACTTTGCATCCCTCTGTTGCTGAG 304
DB 1 Mettysilesersglulemetgluvalleuasnanshvalprophethisglinalgu 20
QY 305 AGTTGGCAATGTGATTTGATTTGTTGGAACCAACCCACATACCTGTAATACACGC 364
DB 21 Sertrpaspasnvalglyleuileuile--glyasnaspylsleuaspillethrglylie 39
QY 365 TTCTGACCAATGACCTGACTGAGAGAGATGATGAGAGGTGCTGCAAAAGAGAGCAGAC 424
DB 40 Leuthrthrleuaspasythrpsaspvalvalaanglnalalegluleuasnhrasn 59
QY 425 CTCATCTCTCTCTACATCCGCTTATCTTCGACCCATGAGCGCATTAACCTGGAACACA 484
DB 60 Thrillelealanhshisproleuilepelysglyvalylsarygillevalaleuaspgly 79
QY 485 TGAAGAGAGCGCTGCTGATCCGGCTCGAGAGACAGAGTGGATATCTCTCTCTCTAT 544
DB 80 Tyrglyser---llelleargylsleuileeglnanasnilleasnleuilealeuans 98
QY 545 ACAGCGTATGATCTGCGCCGCCAGCGCTCAACAACTGGTTGGTAAAGGCTTGAGACT 604
DB 99 Thrasnleuaspvalasnprolysglyvalasnargmetleualaspclnleuglyeu 118
QY 605 TGTACTCTCCAGGCCATCATCTCTCCCAAGCTCCCACTACCTACCTACCA 652
DB 119 GluasnilleasermetileasnhrasnserSerrTYTTLysvalglnthrphelle 138
QY 653 -----GAG 655
DB 139 Prolyasntrillegluasphelysaspserleuasnleuglyleualalysglu 158
QY 656 GGAACCAACCGAGTAGAATTCAACGTTAAC----- 685
DB 159 GlyasnTYrglyutyrCysRphheglusergluglyLysglYlnPhelysProvalgly 178
QY 686 -----TACACCCAAAGCTGACAAAGTCATCTGCACTGAAAGAAATGAC 733
DB 179 AspalaserProtyrilleglYlysleuasp-----Serilleluyryvalasp 194
QY 734 GGTGTTCTGACATCTTTTCTCTGCTAGAGCTGTAATAGAGAACAAACAGCGATTAAT 793
DB 195 Gluilelysleu--gluRhemetllelys--glyasnsglu----- 206
QY 794 CTGAATTTGACTCAGAGGCTTTGATGCAG-----GTGGTAGAT 832
DB 207 leugluilethrglyarglalleleuasnhrasnhrasnprotyrgluthrProvalRheasp 226
QY 833 TTTCCTTCCCGGAACAACACTTTATTCGAGAGAGAAATCTGTGCACTGGAGAGAGCT 892
DB 227 Pheilleysmetasnlysglu----- 233
QY 893 TTGCTTACATACATGAGATGAGAGCGTTATGACACATGATGAATCTGTCTCCGCGCA 952
DB 234 -----serglutrycglyleuclylellelleglYlnleuasnclnthrmetthrleuasp 251

```

```

QY 953 ACCATGATGATCGAATAAAGACACCTAAACTATCTCATATTCGTTAGCCCTGGG 1012
DB 252 GluphesersglutryalalysglinsleuasnhrasnhrasnproserValRgtyr----- 268
QY 1013 GTGGGAGAACCTTGAAGTCTCAATCAAGTCGCGCCCTGTGCTGCTTCTGGAGC 1072
DB 269 Thrlyglinsleu--aspserProilleyslyslvalallellelleglYlyserlylie 287
QY 1073 AGCGTTTCGACGAGGTGTGAG-----GTCGACCTTACCTACCA 1111
DB 288 -----GlyPheglutryrlysalaserGlnleuglyalaspvalRhealthr 303
QY 1112 GGTGAGATGCCATCATGATATCTTGGATGCTGCTCTCCAGAAATAATGCTCATCTC 1171
DB 304 Glyaspillelysleuanshshaspalaleuaspvalalyslleghlnasnvalasnleuasp 323
QY 1172 TGTGACACAGCAACACTGACGAGGCTTTCTTTGACCTTCGAGATATGCTGATTC 1231
DB 324 lleasnhrasn--Tyrserglutryrvalmetlysglyleuylsleuileuileu 342
QY 1232 CACTG-----GAGATTAAGATTAATATATCTATGACACTGACAGGAGCCT 1282
DB 343 TrpleupheylsTYrglinsneInPheProilellyrAlaserGluileasnhrasnpro 362
QY 1283 CTTGAG 1288
DB 363 Pheyls 364

```

## RESULT 5

Q8EP20 PRELIMINARY; PRT; 372 AA.

AC Q8EP20;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
 DE Hypothetical conserved protein.  
 GN OB1940.  
 OS Oceanobacillus thelyensis.  
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.  
 OX NCBI\_TaxID=182710;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HTE831 / DSM 14371 / JCM 11309;  
 RX MEDLINE=22220767; PubMed=12235376;  
 RA Takami H., Takaki Y., Uchiyama I.,  
 RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments.",  
 RL Nucleic Acids Res. 30:3927-3935(2002).  
 DR EMBL: AP004599; BAC13896.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 372 AA; 41869 MW; B66B9C9C8749C43 CRC64;

## Alignment Scores:

Pred. No.:	5.11e-19	Length:	372
Score:	309.50	Matches:	105
Percent Similarity:	46.79%	Conservative:	70
Best Local Similarity:	28.07%	Mismatches:	152
Query Match:	11.10%	Indels:	47
DB:	16	Gaps:	13

US-09-745-506-74 (1-1553) x Q8EP20 (1-372)

```

QY 269 TCCTGAATGACTTTGACCTCTCTGCTGAGAGTTGGAGCAATGTTGATTAAGT 328
DB 14 AlameAspvalTrrpaleprolyshsleualatYrAsPTrpaspasnvalolyleu 33
QY 329 GTGGAACCAAGCCACCACTATCTGTAATACACTCTTCTGACCAATGACTGACTGAG 388
DB 34 lileglyserPheasnhalallethrerlyserlyslvalleu--llethleuaspvalalthr 52
QY 389 GAAGTATGAGAGAGAGTGTGCAAAAGAGCAACCTCATCTCTCTACATCCGCT 448

```

Db	53	SeRvAlValAspArgLualAlaIleGluIbUlyMetAsnLeuIleIleAlaH::SHisSerLeu	72
OY	449	ATPTCCGACCCATGAAGCGCATTAACCTGGCAACACATGGAAGAGCGCCAGGTATCCGG	508
Db	73	LeuHeGInPrOleUlySgInIleAsnMetAsnPrOlySgIArgThValGInIlyS	92
OY	509	GCTGTGGAGAACAGACAGCGGTATCTACTCTCTCTATACAGCCCTTGAATGCTGGCGCCGAC	568
Db	93	LeuIleUlySAsnAPrIleThValTYrAlaSerHisThrAsnLeuAspAlaAsnAsn	112
OY	569	GGCGTCAACAACSTGTGGCTTAAAGCGCCTTGAGCCTTGATCCAGGCCCATACATCCST	628
Db	113	GIYAlaAsnAPrMetLeuAlaThrGluLeuUlySLeuThInSHrThProLeu-----	130
OY	629	TCCAAAGCTCCCAACTACCTACAGAGGAAACCAACCGATGAATTAACSTTAACTAC	688
Db	131	-----ValProValTYr--AsnGluUlySAsnTYrUlyVal-----ValIleTYr	144
OY	689	ACCCAGAC-----CTGACAAAGATGCTACGTGACAGTGAAGAAATGACGGTGTTCCT	742
Db	145	ValProGluSerHisIleGluIbUlySleArgGlnAlaPheAlaGluSerGIYAlaIyTYr	164
OY	743	GTCACCTCTTTTCTGTAGACT-----	766
Db	165	IleGluAsnTYrSerHisCysThrPheGlnSerProGluGlnGIYThrPheUlyProLeu	184
OY	767	-----GCPAAATAG-----GACAAACACGG	787
Db	185	GluGluTYrAspProTYrIleGluIbUlyGlnAsnIlySleThrTYrValAspIbUlyUlyS	204
OY	788	ATTATATCGAATTTGTACTACAGAGGCTTTGATGACAGTGTA--GATTTTCTTTCCCGG	844
Db	205	IleGluThrIleValProGluSerIleUleuSerAsnValIySserIleUleuSerAla	224
OY	845	AACAACAACATTTCAGAAACGGAATTTCTGTCACSTGGAAGAACCTTGGCTCTACAT	904
Db	225	HisProTYrGluGluProAlaTYrAspIleTYrProLeuGluAsnIlyS--ThrThrAsn	243
OY	905	ACTGAATGGAGCGGTATGACACCTGATGACATGTCCTCCCTGCAACCATATTTGAT	964
Db	244	PheGluIleGluArgIleGluAsnLeuThrGluProThrThrLeuAspSerPheIleUlyS	263
OY	965	CGAATTAACAAAGACACTAAACCTATCTATATTCGCTTACGCCCTTGGGGTGGGAGAAC	1024
Db	264	ThrValIlySThrGlnPheGlnLeuGlnIyIleArg-----IleSerGIYAsn	279
OY	1025	TTGAAGTCTACGTCAAAAGTCGTGGCCCTGTGTCTGCTGGTCTTGAGAGACACGTTCTGAG	1084
Db	280	LysAspIlySlySleUlyArgValAlaIleUleuGluGlySerGIYGIUlySerValIleSer	299
OY	1085	GGGTGT-----GAGCGTACCTTTACTCTACACAGGTAGATGTCCCATCATATACT	1133
Db	300	HisAlaMetArgUlySValAsnAPrAllyIleThrIleGlnIlyAsnPrMetThrPheHisIleAla	319
OY	1136	TTTGGATGCTGCTCCCAAGAAATTAATGTCACTCTCTGTGAACACAGACACACTAGAGA	1199
Db	320	GlnAspAlaAlaGluMetGlyLeuThrValIleAspAlaGlnHis--TYrIleUlyS	338
OY	1196	GGCTTTCTTTGTACCTTCGAGATATGCTGGATTTCTCACTTGGAC--ATATAGATAAAT	1255
Db	339	ValMetUlyS TYrTYrThrUlyS GlnIlySLeuAlaGlnThrMetGluLeuAspIlySAspPhe	358
OY	1253	ATTATCTTATCGACACTGACAGGAGCGCTCTTCAGGGTGA	1294
Db	359	IleGluValSerGluIleAsnThrAspProPheGlnPheVal	372

RESULT 6	
08NMWB9	
ID 08NMWB9	PRELIMINARY;
AC 08NMWB9	PRT; 366 AA.
DT 01-OCT-2002	(TREMBL)rel. 22, Created
DT 01-OCT-2002	(TREMBL)rel. 22, Last sequence update)
DT 01-MAR-2003	(TREMBL)rel. 23, Last annotation update)

DE Hypothetical protein MM1511.  
GN MM1511.  
OS Staphylococcus aureus (strain MW2).  
OC Bacteria; Firmicutes; Bacilliales; Staphylococcus.  
OX NCBI\_TaxID=196620;  
|1  
RN  
RP  
RQ SEQUENCE FROM N.A.  
EX MEDLINE=22040717; PubMed=12044378;  
RA Baga T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
RA Yamamoto K., Hiramatsu K.;  
RT "Genome and virulence determinants of high virulence community-  
RT acquired MRSA".  
RL Lancel 359:1819-1827(2002).  
RL EMBL; AP004827; BAB95376.1; -.  
DR InterPro; IPR002678; DUF34.  
DR Pfam; PF01784; DUF34; 1.  
DR TIGRFAMs; TIGR00486; TIGR00486; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 366 AA; 41153 MW; FALF28762D878C79 CRC64;

Alignment Scores:	
Pred. No.:	2.9e-16
Score:	279.00
Percent Similarity:	43.40%
Best Local Similarity:	25.89%
Query Match:	10.00%
Db:	16
US-09-745-506-74 (1-1553) x QBNNM9 (1-365)	
	Length: 366
	Matches: 102
	Conservative: 69
	Mismatches: 151
	Indels: 72
	Gaps: 16

Oy 245 ATGATGATTGAAGGCTCTCCCTTCTTCCTCCGATGACTTGGCATCCCTCGTTTGGAG 304  
 Db 1 MetLysIleAlaSpLeuMetThrLeuLeuAspHisIstValProPheSerThrAlaGlu 20  
 Oy 305 AGTTGGACAATGTTGGATTACTGTGTGGAACCAACCCACACACTGTAATAATACATC 364  
 Db 21 SerThrAspAsnValAlaGlyLeuLeuIleGlyAspGluAspValAlaGluValThrGlyValLeu 40  
 Oy 365 TTCCCTGACCAATGACCTGACGAGACGATGAGAGGTGGCGCAAAAGACAGAC 424  
 Db 41 ---ThrAlaLeuAspCysThrLeuGluValValAsnGluAlaIleGluLysGlyTyrAsn 59  
 Oy 425 CTCATTTCTCTCCATCCATCCGCGCTATCTTCCGACCACTGAAGCGCATACCTGGACACA 484  
 Db 60 ThrIleLeuSerHisIstProLeuIlePheLysGlyValThrSerLeuLysAlaAsnGly 79  
 Oy 485 TGGAAAGAGCGGCTGGTGTATCCGGGCTGCG---GAGAAACAGAGTCGGATATCTACTTCCT 544  
 Db 80 Tyr-----GlyLeuIleIleLeuArgLysLeuIleGlnHisAspIleAsnLeuIleAlaMet 97  
 Oy 542 CATACAGACCTATGATGCTGCGCCCGACGGGCGCTCAACAACATGTTGGCTAAAGGCGCTTGA 601  
 Db 98 HisThrAsnLeuAspAlaAsnProTyrGlyValAsnMetLeuAlaIleAlaIleMetGly 117  
 Oy 602 GCTTTRACCTCCACGGGCCATACATCTCTCCAAAGCTCCCACTACCTCCACA----- 652  
 Db 118 LeuLysAsnIleSerIleIleAsnAsnGlnAsnAspValTyrTyrLysValGlnThrTyr 137  
 Oy 652 ----- 652  
 Db 138 IleProLysAspAsnValAlaGlyProPheLysAspLysLeuSerGluAsnGlyLeuAlaGln 157  
 Oy 653 GAGGGAACCAACCGAGTACATAATC-----AACGTTAACTACACCAACGACACTG 700  
 Db 158 GlnGlyLysAsnTyrIleTyrCysPhePheGluSerGlnGlyArgGlyGlnPheLysProVal 177  
 Oy 701 GACAACATCTATGCTCAGTGAAGAAGATGACGGGTGTTTCTGTCTACTTCTTTTCTGCT 766  
 Db 178 GlyGluAlaAsnProThrIleGlyGlnIleAspLysIleGlyIleTyrVal----- 193  
 Oy 761 AGCAGCTGTAATAGGACAAACACGCGATTAATCTGCAATTGTACTCAGAAAGCGTTTGATG 820

```

Db 194 -----AspGluValLysIleGluPheMetIleAspAlaTyGlnLys----- 207
QY 821 CAGGTGTAGATTTCTTCCCGAACAACAACCTTTATCAGAG----- 865
Db 208 -----SerArgAlaGluGlnLeuIleLysGlnTyHisProTyrGlu 221
QY 866 -----ACGGAATTCCTGTCACTGAGAAACCTTTGCTTACATATCTGAATGGA 916
Db 222 ThrProValPheAspPheIleGluIleLysGlnThrSerLeuTyr-----GlyLeuGly 239
QY 917 CGGTATTCACACATGAGATCTGTCTCCCGCAACCATGATGATGATGAATAAAGA 976
Db 240 ValMetAlaGluValAspAsnGlnMetThrLeuGlnAspPheAlaAlaAspLysSer 259
QY 977 CACCTAAACATATCTCATATTCGCTTACCCCTTGCGGAGAGAACCTTAGAGTCTCAA 1036
Db 260 LysLeuAsnIleProSerValArgPhe-----ValGlyGluSer---AsnGlnLys 275
QY 1037 GTCAAAGCTGGCCCTGTGTGCTGCTGGG-----AGCAGCCTTCTG 1081
Db 276 IleLysArgIleAlaIleIleGlyLysArgIleGlyTyrGlnTyrGlnAlaValGln 295
QY 1082 CAGGCTGTGAGCTGACCTTACCTCAGGTGAGTGCACATCATGATATCTTGGAT 1141
Db 296 GlnGly-----AlaAspValPheValThrGlyAspIleLysHisAspAlaLeuAsp 313
QY 1142 GCTGCTCCCAAGGAATTAATGTCATCTGTGTGAACACACACATGACGAGCTTT 1201
Db 314 ATTCTGATTCCTGAGATATCTGATTCATCTGAGG---AATAAATAATATTTATC 1258
QY 1202 CTTTGTGACCTTCGAGATATCTGATTCATCTGAGG---AATAAATAATATTTATC 1258
Db 333 LysGluGlyLeuLysThrLeuLeuMetAsnArgPheAsnThrGlnLysIleAsnIleAsp 352
QY 1259 CTA-----TCAGAGACTGACAGGAGCCCTCTTCACTGTGTA 1294
Db 353 ValGluAlaSerThrIleAsnThrAspProPheGlnTyrIle 366

RESULT 7
QY 08RE15 PRELIMINARY; PRT; 258 AA.
AC 08RE15;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE NGG1-Interacting factor 3.
GN FN1316.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=1189109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Rosenstein M., Kyrides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010636; AAL95512.1; -.
DR InterPro; IPR002678; DUF34.
DR Pfam; PF01784; DUF34.1.
DR TIGRfams; TIGR00486; TIGR00486.1.
KW Complete proteome.
SQ SEQUENCE 258 AA; 30127 MW; 4A6A2B53B64F4280 CRC64;

```

Alignment Scores: 1.22e-15 Length: 258  
 Pred. No.: 271.50 Matches: 74  
 Score:

```

Percent Similarity: 42.45% Conservative: 61
Best Local Similarity: 23.27% Mismatches: 88
Query Match: 9.73% Indels: 95
DB: 16 Gaps: 7

US-09-745-506-74 (1-1553) x Q8RE15 (1-258)
QY 299 GCTGAGATGGGACAAATGTGGATTTACTGTGTGAGAACCAAGCCACATACCTGTAAT 358
Db 19 AlagIluTUTPrAspAsnIleGlyLeuLeuIleGlyAspTyrAspLysGlu---ValLys 37
QY 359 ACACCTCTCCGACCAATGACCTGACTGAGAGAGTGAAGAGAGCTGCTGCAAAAG 418
Db 38 LysIleGlnPheSerLeuAspAlaThrLeuGlnUserIleGluAsnAlaIleSerGluLys 57
QY 419 GCAGACCTCATCTCTCTTACCATCCGCTTACTTCCGACCATCATGATTAACCTGAG 478
Db 58 ValAspMetLeuIleThrHisHisProIleIlePheLysAlaIleLysAspIleThrGlu 77
QY 479 AACACATGGAAGAGAGCGCTGTGATCCGGGCTGTGAGAACAGAGTGGTATCTACT 538
Db 78 GlnAsnIleLeuGlyLysLysIleArgAsnLeuIleLysAsnAspIleAsnValTyrSer 97
QY 539 CCTCATACAGCTATGATGCTGCGCCGCCAGGCGCTCAACAACATGCTTGAAGGCTT 598
Db 98 IleHisThrAsnLeuAspSerSerIleGluGlyLeuAsnAspTyrValLeuLysLysIle 117
QY 599 GGAGCTGTATCCTCCAGGCCCATCATCTTCCAAAGCTCCCAATACCTTACAGAGGA 658
Db 118 GlnLys----- 118
QY 659 AACCAACGATGAAATTCAGCTTACATACCCCAAGCTGAGCAAAAGTCAATGTCGTA 718
Db 118 ----- 118
QY 719 GTGAAGAATGACGGGTGTTCTGTGCACTTCTTTTCTGTAGAGCTGTAATAGGA 778
Db 119 -----IleSerGluTyrLysIleLeuAspPheAspIleGlu 130
QY 779 CAACACGGAATTAATGCAATTTGATCTCAGAAAGCTTTGATGACAGGTGTATGTTCTT 838
Db 131 Lys-----AsnGly----- 133
QY 839 TCCCGGACAAACAACATTTATCAGAAAGCAGAAATTTGTCTACGTGAGAAAGCTTGCCT 898
Db 133 ----- 133
QY 899 CTACATCTGAGATGAGGAGGTTATGACACACTGATGATCTGTCTCCCTGCAACCATG 958
Db 134 -----GlyIleGlyArgIlePheLysLeuAsnGlnGluLysAsnLeuLysPhe 150
QY 959 ATTGATCGAATTAAGAAAGACACCTAAACATATCTCATATTCGCTTACCCCTGGGTGGG 1018
Db 151 IleGluGluLeuLysLeuLysLeuLysIleLeuAsnLeuArgVal-----IleSer 167
QY 1019 AGACCTTAGACTCTCAAGTCAAGTCAAGTGGGCCCTGTGCTGCTGTGAGAGCAGCTT 1078
Db 168 AsnAspLeuAsnLysLysIleLysLysValAlaIleAlaIleAsnGlnIleAlaMetAsnTyr 187
QY 1079 CTGCAAGGGGT-----GAGGCTGACCTTACCTACACAGAGGAGATGTCACATCT 1129
Db 188 TrpLysLysAlaLysLysGluLysValAspLeuPheIleThrGlyAspValSerTyrHis 207
QY 1130 GATACCTTGGATGCTGCTTCCCAAGGAATTAATGTCATCTCTGGAACAGCAACACT 1189
Db 208 AspAlaLeuAspAlaLeuGlnLysnGlyLeuSerValIleAspPheGlyHisTyrGluSer 227
QY 1190 GAGCAGGCTTCTTCTGACCTT---CGAGATATCTGAGATTCATCTGAGAG 1240
Db 228 GluTyrPhePheTyrGluIleLeuIleGluGluLeuLysAspAsnAsnLeuGlu 245

RESULT 8
Q8SV52

```







```

QY 653 GAGGAACGACGAGTACGATTCACGTTAACTACACCCAGACGCTG-----GACAAA 706
    ||| ||| : : : : : ||| : : : : : ||
Db 139 TPGGlyValHisValIleuProIysAspAlaIatYtIleuYsMetIlePheAspAla 158
QY 707 GTCATGCTGTCAGTGAAGA-----ATTGACGCTGTTCT--- 742
    ||| : : : : : ||| : : : : : ||
Db 159 GtAlaGtAlaIleIleIysPrtYtArgGtIuCsYsAlaPheGtIuIleGtIuYtHtGtIu 178
QY 743 -----GTCACCTCTTTTCTGCTAGAGCTGGTAATGAGACAACACGGAATTAAT 793
    ||| : : : : : ||| : : : : : ||
Db 179 PheArgProValGtIuGtAlaAsnProAlaGtIuYsAspAlaPlys----- 194
QY 794 CTGAATTTGTAACGAGAGCTTTGATGTCAGAGCTGATTTCTTCCCGGACAACAA 853
    ||| : : : : : ||| : : : : : ||
Db 195 -----LeuPheYsSerIleuGtIuIleGtIuPheValAlaProAlaGtIu 211
QY 854 CTTTATCAGAGACGGAATTTCTG-----TCACTGAGAGACCTTTGCTT--- 898
    ||| : : : : : ||| : : : : : ||
Db 212 ArgAlaArgLeuHtHtSerValIleuArgGtIuAlaHisProTyrGtIuIuProAlaPheAsp 231
QY 899 -----CTACAT-----ACTGGAATGGAGCGTTATGC 925
    ||| : : : : : ||| : : : : : ||
Db 232 IleValGtMetHisSerAlaGtIuSerIleuGtIuAsnAlaHtHtGtIuGtIuArgValGtIu 251
QY 926 ACACGATGATGATCTGCTCCCTGCGACACCATGATGATGATTAAGACACGCTAAA 985
    ||| ||| : : : : : ||| : : : : : ||
Db 252 GtIuLeuProGtIuProMetArgIleuAlaAspPheValGtIuGtIuAlaAsnAlaPhePro 271
QY 986 CTAATCATATTCGCTTACGCTTGGGCTGGGAGAAC-----TTAGAGTCAAGTC 1039
    ||| : : : : : ||| : : : : : ||
Db 272 ValHtGtIu-----TtGtIuValArgAlaHtHtGtIuArgProGtIuGtIuMetVal 287
QY 1040 AAAGTGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1090
    ||| : : : : : ||| : : : : : ||
Db 288 SerArgValAlaValIleSerGtIuSerGtIuYsSerPheIleuAsnAspValIleYsLeu 307
QY 1091 GAGGCTGACCTTATCTCAGACAGTCCAGATGCTCCATGATGATCTTGTGATGCTGTCC 1150
    ||| : : : : : ||| : : : : : ||
Db 308 GtYValAspValIuYtValHtHtSerAspIleuArgHisIleProValAspGtIuYtIleuArg 327
QY 1151 CAA-----GGAATTAATGATCTCTCTGTGAACACAGACACAGGCTTTCTTTCT 1207
    ||| ||| : : : : : ||| : : : : : ||
Db 328 GtIuGtIuYtProAlaValIleAspHtAlaHisIleTtPheIleAspIlePhePrtHtSer 347
QY 1208 GACCTTGAATGATGCTG-----GATTCACCTTGAATGAATGAATGAATGAATGA 1264
    ||| : : : : : ||| : : : : : ||
Db 348 GtAlaGtIuGtIuIleuGtIuAspIlyAlaProGtIuValGtIuValIleSerIle 367
QY 1265 GAGACTGAC 1273
    ||| |||
Db 368 ArgTtHtAsp 370

```

```

RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.L.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rooke P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.,
RA "Genome of the bacterium Streptococcus pneumoniae strain R6.",
RA J. Bacteriol. 183:5709-5717(2001).
RA EMBL: AE008515: AML00266.1: -.
RA Hypothetical protein; Complete proteome.
SQ SEQUENCE 265 AA; 29871 MW; 4939542E19C68826 CRC64;

Alignment Scores:
Pred. No.: 1,72e-09 Length: 265
Score: 203.50 Matches: 66
Percent Similarity: 36.688 Conservative: 62
Best Local Similarity: 18.918 Mismatches: 124
Query Match: 7,308 Gaps: 97
DB: 16 Indels: 7

US-09-745-506-74 (1-1553) x QBDN4 (1-265)
QY 260 CTCTCTTCTCTCTTGAATGACCTTTCATCCCTCTCTCTCTGAGACTGGACATGTT 319
    ||| : : : : : ||| : : : : : ||
Db 6 ValIleGtIuAlaTyrGtIuAlaPheCysProGtIuPheSerMetGtIuYsAspSerCys 25
QY 320 GGAATTAAGTGGAGAACCCAGCCACACATAGTGAATGAATGACCTTCTGACCAATGAC 379
    ||| ||| : : : : : ||| : : : : : ||
Db 26 GtYLeuGtIuIle---GtYtHtLeuAspIlyGtIyIleIleHtArgValMetValIleAsp 44
QY 380 CTACAGTGAAGATGATGAGAGAGTGTCTCAAAAGAGAGACACGCTCATCTCTCTAC 439
    ||| ||| : : : : : ||| : : : : : ||
Db 45 IleArgGtIuHtHtValAlaIleGtIuAlaIleGtIuIleGtIuYsAlaPheIleIleValIys 64
QY 440 CATCCGCTATCTTCCGACCCATGAAGCGCATACCTGGAAACATGGAAGAGCGCTG 499
    ||| ||| ||| : : : : : ||| : : : : : ||
Db 65 HisAlaProIlePheArgProIleYsAspIleuAlaSerArgProGtIuAsnGtIuIle 84
QY 500 GTATCGGCGCTGTGAGACAGAGTCCGATCTACTCTCTATACAGCTGATGATGCT 559
    ||| ||| : : : : : ||| : : : : : ||
Db 85 TyrIleAspIleuIleYsHisAspIleAlaValIyTValSerHisIleAspIle 104
QY 560 GCGCGCCAGGCGCTCAACAACGTGGTGAAGCGCTTGAGCTGTACCTCCAGGCC 619
    ||| ||| : : : : : ||| : : : : : ||
Db 105 ValGtIuAsnGtIuYsAsnAspTtPheCysGtIuMetLeu----- 117
QY 620 ATGACATCTCTCAAAAGTCCCACTACCTACAGAGGAGAAACACAGAGTGAATTCAC 679
    ||| : : : : : ||| : : : : : ||
Db 117 ----- 117
QY 680 GTTAACATACACCCAGACGCTGACAAAGTCATGCTGACGTAAGGAATGACGCTGT 739
    ||| : : : : : ||| : : : : : ||
Db 118 -----GtYIle 119
QY 740 TGTGTCACTTCTTTTCTGCTAGAGCTGTAATGAGAGACAACAGGATTAATGTAAT 799
    ||| : : : : : ||| : : : : : ||
Db 120 GtIuGtIuHtHtYtIleuGtIuHtGtIuProGtIuArg----- 132
QY 800 TGTACATCAGAGGCTTGTGATGACAGGTGTGATTTCTTCCGGAACAACAATTAT 859
    ||| : : : : : ||| : : : : : ||
Db 132 ----- 132
QY 860 CAGAGACGGAATTTCTGTCATGAGAGAACCTTTGCTCTACATAGTGAATGAGAGG 919
    ||| : : : : : ||| : : : : : ||
Db 133 -----GtYIleGtYArg 136
QY 920 TTATGCACTGATGATGATCTGTCTCCCTGCAACATGATGATGATGAATGAAGACAC 979
    ||| : : : : : ||| : : : : : ||
Db 137 IleGtYAsnIle---GtIuProGtIuHtPheTtPgluIleuAlaGtIuValIysGtIuVal 155
QY 980 CTAACATATCTATATGCTTGAAGCCCTTGGGCTGGGAGAACCTTAGAGTGCAGTGC 1039
    ||| ||| : : : : : ||| : : : : : ||
Db 156 PheAspIleuAspSerIleuArgMetValHisTyrGtIuGtIuAsnAspIleuGtIuProIle 175

```

```

QY 1040 AAGTCGCGCCCTGTGCTGCTTCTTGGAAGACGCTTCGAGGGTGT----- 1090
DB 176 SerArgValAlaIleCysGlySerGlnSerPheTyrLysAspAlaLeuAlaLys 195
QY 1091 GAGCGCTACCTTTCACCTCAGAGTGAGTGCCTCATGATCTTGTATGCTGCTTC 1150
DB 196 GlyAlaAspValTyrIleThrGlyAspIleTyrThrAlaGlnAspMetLeuSer 215
QY 1151 CAGGAATAAATGTCATCTCTGTGACACAGCAACACTGACGAGCTTCTTCTGAC 1210
DB 216 AspGlyLeuAlaLeuAspProGlyHis--TyrIleGluValLeuPheValGluLys 234
QY 1211 CTTCGAGATATGCTGATCTTCACCTGAGATATAG-----ATTAATATTTCCTA 1261
DB 235 IleAlaAlaLeuLeuThrGlnTyrPlyGlnLysGlyTyrGluLeuGluIleuPro 254
QY 1262 TCAGAGACTGACAGGACCTCTTCAG 1288
DB 255 SerGlnAlaSerThrAsnProPheArg 263

```

## RESULT 12

```

Q8G5M5 PRELIMINARY; PRT; 309 AA.
ID O8G5M5
AC O8G5M5
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein with duf34.
GN BL0984.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karimiantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,
RA Fridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL, AE014721; AAN24792.1;-.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 309 AA; 33851 MM; 683448905FP0F3 CRC64;

```

## Alignment Scores:

```

Pred. No.: 1,81e-09 Length: 309
Score: 203.50 Matches: 67
Percent Similarity: 36.48% Conservative: 45
Best Local Similarity: 21.82% Mismatches: 106
Query Match: 7.30% Indels: 89
DB: 16 Gaps: 5

```

US-09-745-506-74 (1-1553) x Q8G5M5 (1-309)

```

QY 248 GATTGAAGCTCTCCCTTCTCTCTGAATGACTTTGACCTCCCTCTGCTGAGAGT 307
DB 6 AsnLeuLysGlnValIleAspValLeuGlnThrLeuTyrProLeuArgTyrAlaGluLys 25
QY 308 TGGGACATGTTGATTAAGTGGAGACCAAGCCACACATACGTAAATACATCTTC 367
DB 26 TrpAspGluProGlyLeuIleVal--GlyAspLeuSerHisAspValHisArgIleVal 44
QY 368 CTGACCAATGACCTGACGAGAGAGTGAAGAGTGTGCAGAAAGAGCAGACCTC 427
DB 45 PheAlaAlaAspProThrSerAlaIleLeuAspLysAlaIleAlaThrGlyAlaAspLeu 64
QY 428 ATTCTCTCTACCATCGCGCTATCTTCGACCCCATGAAGCCCATACCTGGAACATG 487
DB 65 LeuIleThrHisHisProLeuPhePheArgSerValHisGluThrSerGlyLeuGlyPhe 84

```

```

QY 488 AAGAGCCCTGTGATCCGGCTCTGAGAAACAGATCGGTATCTACTTCATACCA 547
DB 85 ArgGlyAspIleValAlaArgLeuTyrGlnHisGlyCysGlyLeuTyrValGlyHisThr 104
QY 548 GCGTATGATGCGGCGCCAGGCGCTCAACACTGCTTGGCTTAAGGCTTGACCTGT 607
DB 105 AsnAlaAspAlaIleTyrArgGlyValGlyAlaIleAlaIleAspTyrPheGlyLeuIle 124
QY 608 ACCTCCAGCCCATCATCTTCACCAAGCTCCACATACCTCTACAGAGGAACCCAGA 667
DB 125 AspGlnLysProLeuValProIleAspAspAlaAsn-----AlaSerHisPro 140
QY 668 GTACATTTCAACGTTAACTACACCAAGACCTGGCAAGATCATCTGCAAGTAAAGA 727
DB 141 Val----- 141
QY 728 ATTGACGGTGTTCGTACATCTTTTCTGCTAGACTGGTAAATAGAGAAACAACAGG 787
DB 141 ----- 141
QY 788 ATTAATCTGAATTGACTACAGAGGCTTTGATGAGTGATGATTTTCTTCCGGAAC 847
DB 141 ----- 141
QY 848 AAACACTTTATCAGAAAGCGAAATTCTGTCACCTGAGAAAGCTTCTCTACATACT 907
DB 141 ----- 141
QY 908 GGAATGGAGCGGTTATGACACATGATGATCTGCTCCGCAACCATGATGATGCA 967
DB 142 GlyLeuGlyArgValGlyArgLeuProLysProValAlaIleLeuLysAspPheAlaGlnArg 161
QY 968 ATAAAGACACCTAAACATATCTCATTCGCTTAGCCCTGGGGTGGGGAACCTTA 1027
DB 162 ValPheAspGluValSerAspHisGlyMetThrSerAlaLeuGlyIleGlnValCysGly 181
QY 1028 GAGCTCTCA-----GTCAAGTCTGGCCCTGTGCTGCTCGGAGACAGCTTCTG 1081
DB 182 AspSerAspThrPheValAlaGlnTyrValAlaIleLeuProGlySerGlyAspSerLeuPhe 201
QY 1082 CAGGCTGTGAGGCT-----GACCTTATACCTCAGAGGAGATGCTCCATCATGAT 1132
DB 202 AsnGluValAlaArgAlaThrGlyValAspValIlyValIThrSerAspLeuArgHisHisPro 221
QY 1133 ACTTGTGATGCTGCTTCCCA 1153
DB 222 ValThrAspAlaIleGluGln 228

```

## RESULT 13

```

Q8DT99 PRELIMINARY; PRT; 262 AA.
ID Q8DT99
AC Q8DT99;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SMU.1463C.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=DA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., Mcshan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans DA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL, AE014978; AAN59121.1;-.
KW Hypothetical protein; Complete proteome.

```

SQ SEQUENCE 262 AA: 29740 MW: A5740C55B8F248B CRC64;  
 Alignment Scores:  
 Pred. No.: 3,15e-08 Length: 262  
 Score: 189.50 Matches: 66  
 Percent Similarity: 36.59% Conservative: 50  
 Best Local Similarity: 20.82% Mismatches: 100  
 Query Match: 6.79% Indels: 101  
 DB: 16 Gaps: 10  
 US-09-745-506-74 (1-1553) x Q8D799 (1-262)  
 QY 353 GTAATACACTCTTCGACCAATGACCTGAGAGAGTGTGAGGAGAGGAGTGTGCA 412  
 Db 36 lIeLysArgLeuMetIleAlaLeuAspValArgIuThrThValAlaGluAlaIleGlu 55  
 QY 413 AAGAAGGACGACCTATCTCTCTCTACCACTCCGCTATCTTCGACCCATGAGAGCGCAT 472  
 Db 56 LysLysValAspLeuLeuIleValLysHisAlaProIlePheArgProLeuLysAsnLeu 75  
 QY 473 ACCTGGAAACATGGAAGGAGCGCTGATCCGCGCTGGAGAGAGAGTGGTATC 532  
 Db 76 Val---GluThrAlaGlnAsnHisIleTyPheAsnLeuIleLysHisAspIleAlaVal 94  
 QY 533 TACTCTCTCATACAGCCTATGATGCTGCGCCGACGAGGCTCAACAAGCTGTGGCTAA 592  
 Db 95 TyValSerHisThraSnIleAspIleValProAspGlyLeuAsnSprTrPheCysAsp 114  
 QY 593 GGGCTTGGAGCTTGATCTCCAGGCCATACATCTCTCCAAAGCTCCCAACTACCTTACA 652  
 Db 115 LeuLeuAspIleLysAsnArgArgIleLeuSerProSerLysAsp----- 129  
 QY 653 GAGGAAACACCGAGTAGAATTCAAGCTTAACACACCAAGACCTGGACAAAGTCATG 712  
 Db 129 ----- 129  
 QY 713 TCTGACGTAAAGAAATGACGGTGTCTGTCACTCTTTCTGTAGAGACTGTAAAT 772  
 Db 129 ----- 129  
 QY 773 GAGGAACAAACAGGATTAACTGAAATGTACTCAAGAGCTTGTGATGAGTGTAGAT 832  
 Db 130 ----- 130  
 QY 833 TTTCTTTCCCGAACAACAACCTTATCAGAAAGCGAAATCTGTCACTGAGAGAGCCT 892  
 Db 131 Tyr----- 131  
 QY 893 TTGCTTTCTACATACCTGAAATGGAGCGTATATGACACTGGATGATCTGTCCCTGGCA 952  
 Db 132 -----GlyIleGlyArgValGlyAspIle---SerProLeuSerPheGlu 145  
 QY 953 ACCATGATGATGCAATTAAGAAACACCTTAACATTCATATTTGGCTTA---GCCCTT 1009  
 Db 146 AspLeuAlaLysLysValLysLysIlePheAsnLeuAspSerValArgLeuValSerTy 165  
 QY 1010 GGGGTGGGAGAAAGCTTACAGTCAAGTCAAGAGTGGCTGCTGCTGGTGGG 1069  
 Db 166 GlyGluAsnAsnProLeuIleSerArg-----ValAlaIleCysGlyGlySerGly 182  
 QY 1070 AGCAGCGTCTGAGAGGTGT-----GAGCTGACCTTTACCTCAACAGTGTGATG 1120  
 Db 183 GlnSerPheTyGlnIuAlaLeuThrLysGlyAlaGlnValTyrlIleThrGlyAspIle 202  
 QY 1121 TCCCATCATGATATCTTGGATGCTGCTCCCAAGAAATTAATGATCTCTGTGAACAC 1180  
 Db 203 TyTyThHisThraLaglnGlnIleuIleuThraSnGlyLeuLeuAlaLeuAspPheGlyHis 222  
 QY 1181 AGCAACAGTGAAGAGGCTTCTTCTGACCTCGAGATATGCTGAT-----TCTCAC 1234  
 Db 223 ---HisIleGluValLeuPheValArgGluLeuAlaGlnLysPheGlnThrTrpSerCys 241  
 QY 1235 TTGGAGAAAT---AAGATTAATATTATTCATCAGAGACTGACAGGAGCCT 1282

Db 242 GlnGluAsnTrpAspIleThrIleLeuGlnSerGlnValAsnThrAsnPro 258  
 RESULT 14  
 ID P95777 PRELIMINARY; PRT; 262 AA.  
 AC P95777;  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE ORF4 protein.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=XC;  
 RX MEDLINE=97175539; PubMed=9023194;  
 RA Tsukitoka Y., Yamashita Y., Oho T., Nakano Y., Koga T.;  
 RT "Biological function of the dtdp-ribose synthesis pathway in  
 RT Streptococcus mutans.";  
 RL J. Bacteriol. 179:1126-1134(1997).  
 DR EMBL: D78182; BAA11246.1;  
 DR InterPro: IPR002678; DUF34.  
 DR Pfam: PF01784; DUF34; 1.  
 DR TIGRDB: TIGR00486; TIGR00486; 1.  
 SQ SEQUENCE 262 AA: 29723 MW: 68E0DBBA6002DE3 CRC64;  
 Alignment Scores:  
 Pred. No.: 7.23e-08 Length: 262  
 Score: 185.50 Matches: 65  
 Percent Similarity: 36.28% Conservative: 50  
 Best Local Similarity: 20.50% Mismatches: 101  
 Query Match: 6.65% Indels: 101  
 DB: 2 Gaps: 10  
 US-09-745-506-74 (1-1553) x P95777 (1-262)  
 QY 353 GTAATACACTCTTCGACCAATGACCTGAGAGAGTGTGAGGAGAGGAGTGTGCA 412  
 Db 36 lIeLysArgLeuMetIleAlaLeuAspValArgIuThrThValAlaGluAlaIleGlu 55  
 QY 413 AAGAAGGACGACCTATCTCTCTCTACCACTCCGCTATCTTCGACCCATGAGAGCGCAT 472  
 Db 56 LysLysValAspLeuLeuIleValLysHisAlaProIlePheArgProLeuLysAsnLeu 75  
 QY 473 ACCTGGAAACATGGAAGGAGCGCTGATCCGCGCTGGAGAGAGAGTGGTATC 532  
 Db 76 Val---GluThrAlaGlnAsnHisIleTyPheAsnLeuIleLysHisAspIleAlaVal 94  
 QY 533 TACTCTCTCATACAGCCTATGATGCTGCGCCGACGAGGCTCAACAAGCTGTGGCTAA 592  
 Db 95 TyValSerHisThraSnIleAspIleValProAspGlyLeuAsnSprTrPheCysAsp 114  
 QY 593 GGGCTTGGAGCTTGATCTCCAGGCCATACATCTCTCCAAAGCTCCCAACTACCTTACA 652  
 Db 115 LeuLeuAspIleLysAsnArgArgIleLeuSerProSerLysAsp----- 129  
 QY 653 GAGGAAACACCGAGTAGAATTCAAGCTTAACACACCAAGACCTGGACAAAGTCATG 712  
 Db 129 ----- 129  
 QY 713 TCTGACGTAAAGAAATGACGGTGTCTGTCACTCTTTCTGTAGAGACTGTAAAT 772  
 Db 129 ----- 129  
 QY 773 GAGGAACAAACAGGATTAACTGAAATGTACTCAAGAGCTTGTGATGAGTGTAGAT 832  
 Db 130 ----- 130  
 QY 833 TTTCTTTCCCGAACAACAACCTTATCAGAAAGCGAAATCTGTCACTGAGAGAGCCT 892

```

Db      131 Tyr-----
QY      893 TTGGTTTACATGATGAGGAGCGATTATGACACATGATGATGCTGCTCCCTGCA 952
Db      132 -----GlyIleGlyArgValGlyAspIle-----SerProLeuSerHeu 145
QY      953 ACCATGATGATTCGAATATAAAGACACCTAAACCTATGTCATATGCTTA---GCCCTT 1009
Db      146 AspleuAlaLysValLysValLysIleHeuAsnLeuAspSerValArgLeuValSerTyr 165
QY      1010 GGGGGGGGAGAACCTTTCAGCTTCACATCAAGTCAGCGCCCTGCGCTGCGCTGCGG 1069
Db      166 GlyIleuAsnAspProLeuIleSerArg-----IleAlaIleCysGlyLeuSerGly 182
QY      1070 AGCAGCGCTTCGACGGGTGT-----GAGCGTGCACCTTTCACCTACAGGTGAGATG 1120
Db      183 GlnSerPheTyrGlnGlnAlaLeuThrLysGlyAlaGlnValTyrIleThrGlyAspIle 202
QY      1121 TCCCATGCTGATCTTGGGATGCTGCTCCCAAGGATTAATGTCATCCTCTGGAACAC 1180
Db      203 TyrTyrHisThrAlaGlnGlnMetLeuThrAsnGlyLeuAlaLeuAspProGlyHis 222
QY      1181 AGCAACACTGACGAGCGCTCTTCTGACCTTCGAGATATGCTGAT-----CTCACC 1234
Db      223 ---HisIleGluValLeuPheValArgLysLeuAlaGlnLysPheGlnThrTyrSerGys 241
QY      1235 TTGGAGAAAT---AAGATTAATATTTATCTATCAGAGACTGACAGGACCT 1282
Db      242 GlnGluAsnTyrPaspIleThrIleLeuGlnSerGlnValAsnThrAsnPro 258

```

## RESULT 15

```

Q9LAG2 PRELIMINARY; PRT; 265 AA.
AC
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical 29.8 kDa protein.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=53/7;
RX MEDLINE=20120540; PubMed=10653753;
RA Savijoki K., Palva A.;
RT "Purification and molecular characterization of a tripeptidase from
   Lactobacillus helveticus";
RL Appl. Environ. Microbiol. 66:794-800(2000).
DR EMBL, AJ243321; CAB72937.1; -.
DR InterPro: IPR002678; D0F34.
DR Pfam: PF01784; D0F34. 1.
DR TIGRFAMs: TIGR00486; TIGR00486; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 29850 MW; 716C9765F4860D17 CRC64;

```

## Alignment Scores:

```

Pred. No.: 1.67e-07 Length: 265
Score: 181.50 Matches: 76
Percent Similarity: 37.32% Conservative: 55
Best Local Similarity: 21.65% Mismatches: 123
Query Match: 6.51% Indels: 97
DB: 2 Gaps: 12

```

US-09-745-506-74 (1-1553) x Q9LAG2 (1-265)

```

QY      251 TTGAAGGCTTCCTTCTTCCTGATGATCTTGCATCCCTCTCTTCTGAGAGATTGG 310
Db      4 ValLysAspIleValAsnArgLeuArgLysPheProGluAspIleAlaSerLysGly 23
QY      311 GACATGTTGGATTACTGTGGAGAACCAAGCCACACACATACGTAAATACACTCTTCCTG 370

```

```

Db      24 AspProValGlyMetGlnIle---GlySerMetGlnAlaAspValThrLysValMetThr 42
QY      371 ACCAATGACCTGATGAGAGATGATGAGAGGCTGTCGCAAAAGAGACACTGAT 430
Db      43 ThrLeuAspValArgProGlnValValGlnGlnAlaValGlnLysGlyValAsnPheIle 62
QY      431 CTCCTCCATCACTCCGCTATCTCTCCGACCATGAAAGCCCATPACCTGGAACATGAG 490
Db      63 ValSerHisHisProValMetPheArgProAlaArgAsnLeuAspPheAlaAspAlaGln 82
QY      491 GAGCCCGCTGGATGATCCGGGCTCTGGAGAACAGAGTGGTATCTACTCTCCATPACCC 550
Db      83 AsnAlaMetTyrGlyAsnIleLeuLysAsnGlyIleThrValTyrSerIleHisThrAsn 102
QY      551 TATGATGCTGGCCCGCCAGGCGTCAACACTGCTGGCTTAAGAGGCTTGGAGCTTACC 610
Db      103 SerAspLysAlaGlnAspGlySerLysAspIleProGlnAlaGlnGlnLeuGlnLysAsp 122
QY      611 TCCAGCCCATPACATCTTCCAAAGCTCCCAACTACCTACAGAGGAGAACCCAGAGTA 670
Db      123 ValGluProPhe-----
QY      671 GAATTCACAGTTAACTACACCCAGACCTGGAGAACAGTCATGTCGACGTGAAGAAAT 730
Db      127 -----CysLeuAspAsp-----
QY      731 GACGGTGTCTGTCATCTTCTTCTGCTAGACTGTAATGAGAACAAACAGGAT 790
Db      131 AspGlyIleAlaMet-----GlyArgLysGlyArgLeuProGlnThr----- 144
QY      791 AATCTGAATTGACTACAGAGGCTTGAATGCAAGGTGTAGATTTCTTCCCGGACAA 850
Db      145 -----MetThrAlaTyrAspPhe-----
QY      851 CAACCTTACGAAAGACGGAATTCGTACAGGAGAGACCTTCTCTACATACGTGA 910
Db      151 ---AlaTyrTyrValLysGlnAsnMetGlyIleLys-----
QY      911 ATGGAGCGGTATGACACACTGATGAATCTGTCTCCCTGGCAACCATGATTGATCGAATA 970
Db      162 MetaIaArgLeuIleThrAlaAspAsnAsnLysLysIleSerThr----- 176
QY      971 AAAAGACACCTAAACCTATTCATATTCGCTTAGCCCTTGGGGTGGGAGAACCTTAG 1030
Db      176 -----
QY      1031 TCTCAAGTCAAAAGTCGTGGCCCTGTGTGCTGTCTGGAGAGCGTTCTGCAGGGTGT 1090
Db      177 -----ValGlyPheIleCysGlyAspGlyLysTyrThrArgAlaLeu 192
QY      1091 GAG-----GCTGACCTTTACCTCACAGGTGAGATGCCATCATGATACTTTGGATGCT 1144
Db      193 AspAspHisLeuAspAlaPheIleThrGlyAspValTyrTyrHisValGlyHisAspMet 212
QY      1145 GCTTCCCAAGAAATAATGATCTCTGTGAACACAGACCAACAGACGAGGCGCTTCT 1204
Db      213 IleSerSerGlyLeuThrValValAspProGlyHis---TyrThrGlnLysLeuPheLys 231
QY      1205 TCTGACCTTCGAGATGCTGATGCT---CACTTGAGAAAT-----AAGATAAATATT 1255
Db      232 TyrLysValTyrAspArgLeuLysLysTyrPasnGlnGlnAsnAsnTyrPasnValGlyVal 251
QY      1256 ATCTTATCAGAGACTGACAGGAGCCCTTTCAG 1288
Db      252 GluLeuSerGlnValSerThrAsnProPheGln 262

```

Search completed: August 22, 2003, 14:05:51  
Job time: 174 secs